Università di Genova

DIBRIS DIPARTIMENTO
DI INFORMATICA, BIOINGEGNERIA,
ROBOTICA E INGEGNERIA DEI SISTEMI

Apprendimento di SQL basato su errori frequenti

Giovanna Guerrini
& Davide Ponzini

[ITADINFO]

CONVEGNO ITALIANO
SULLA DIDATTICA DELL'INFORMATICA

PADOVA

9-11 OTTOBRE 2026

Materiale del laboratorio

https://bit.ly/itadinfo26-sql



Contenuti

1. Errori frequenti e loro classificazione
2. Cosa può aiutare?
   a) cause (misconcezioni)
   b) analogie, visualizzazioni, tool
3. E l'IA generativa?
   a) spiegazione di errori
   b) generazione di esercizi ad hoc

Ma prima ...

wooclap.com/UKUOEZQ



Errori frequenti

Errori frequenti

- Schema adattato dalla letteratura [Miedema et al., 2021]
- Gli attributi sottolineati formano le chiavi

Tabelle	Attributi
customer	cID, cName, street, city, age
store	sID, sName, street, city
product	pID, pName, suffix
inventory	sID, pID, date, quantity, unit_price

D. Miedema, E. Aivaloglou, G. Fletcher. Identifying SQL Misconceptions of Novices: Findings from a Think-Aloud Study. In Proc. ACM ICER 2021, pp. 355–367. https://doi.org/10.1145/3446871.3469759

Esempi di errori frequenti

Tabelle	Attributi
customer	cID, cName, street, city, age
store	sID, sName, street, city
product	pID, pName, suffix
inventory	sID, pID, date, quantity, unit_price

SELECT WHERE *

Sintattico, su qualsiasi schema

SELECT * FROM customer WHERE name = alice

Sintattico, su questo schema

SELECT * FROM customer WHERE cName = ‘alice’

Esempi di errori frequenti

Tabelle	Attributi
customer	cID, cName, street, city, age
store	sID, sName, street, city
product	pID, pName, suffix
inventory	sID, pID, date, quantity, unit_price

SELECT * FROM customer WHERE ‘alice’ = ‘bob’

SELECT * FROM customer
WHERE cName = ‘alice’ AND cName = ‘bob’

Semantici (sbagliato qualsiasi sia la richiesta)

Esempi di errori frequenti

Tabelle	Attributi
customer	cID, cName, street, city, age
store	sID, sName, street, city
product	pID, pName, suffix
inventory	sID, pID, date, quantity, unit_price

I clienti il cui nome è alice

SELECT * FROM customer

WHERE city = ‘alice’

Logici
(non corrispondono alla richiesta)

I codici dei negozi che non hanno prodotti che costano più di 200 Euro in inventario

SELECT DISTINCT sID FROM inventory

WHERE unit_price <= 200

Esempi di errori frequenti

Tabelle	Attributi
customer	cID, cName, street, city, age
store	sID, sName, street, city
product	pID, pName, suffix
inventory	sID, pID, date, quantity, unit_price

I codici dei negozi che hanno prodotti che costano più di 200 Euro in inventario

```
SELECT DISTINCT sID
FROM inventory NATURAL JOIN product
NATURAL JOIN store
WHERE unit_price <= 200
```

Complicazioni

Le città in cui si trovano negozi, senza ripetizioni

```
SELECT city FROM store GROUP BY city
```

Errori frequenti – una tassonomia

[SYS] Syntactic errors

[LOG] Logical errors

[SEM] Semantic errors

[COM] Complications

- https://github.com/DavidePonzini/sql_error_taxonomy/tree/revision

Toni Taipalus, Mikko Siponen, and Tero Vartiainen. 2018. *Errors and Complications in SQL Query Formulation.* ACM Trans. Comput. Educ. 18, 3, Article 15 (September 2018), 29 pages. https://doi.org/10.1145/3231712
Davide Ponzini, Barbara Catania, Giovanna Guerrini. *A Proposal for Revising SQL Error Taxonomies Based onAutomated Detection*. DataEd 2026

Possibili cause e possibili rimedi

Un esempio

Numero di negozi per città

```
SELECT city, COUNT(DISTINCT city)
FROM store
GROUP BY city;
```

Possibili cause?

Esercizi per chiarire?

Tabelle	Attributi
customer	cID, cName, street, city, age
store	sID, sName, street, city
product	pID, pName, suffix
inventory	sID, pID, date, quantity, unit_price

Analogia con mondo reale?

Visualizzazione?

Strumenti?

Un esempio

Tabelle	Attributi
customer	cID, cName, street, city, age
store	sID, sName, street, city
product	pID, pName, suffix
inventory	sID, pID, date, quantity, unit_price

Numero di negozi per città

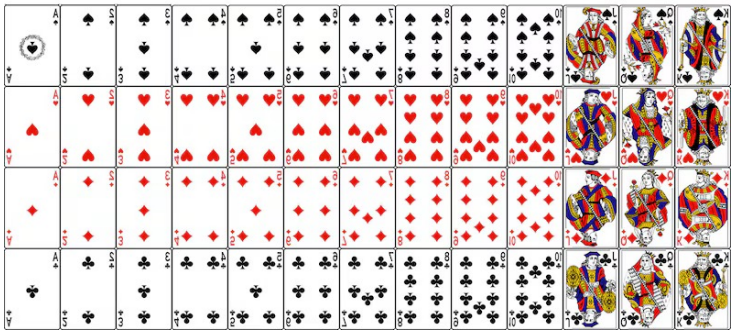
```
SELECT city, COUNT(DISTINCT city)
FROM store
GROUP BY city;
```

Visualizzazione?

city	sID	SName	street
Amsterdam	0	Coop	…
	…	…	…
Utrecht	1	Lidl	…
	…	…	…
Eindhoven	2	Lidl	…
	…	…	…
Breda	3	Hoogvliet	…
	4	Sigro	…
	…	…	…

Possibili cause?

Analogia con mondo reale?

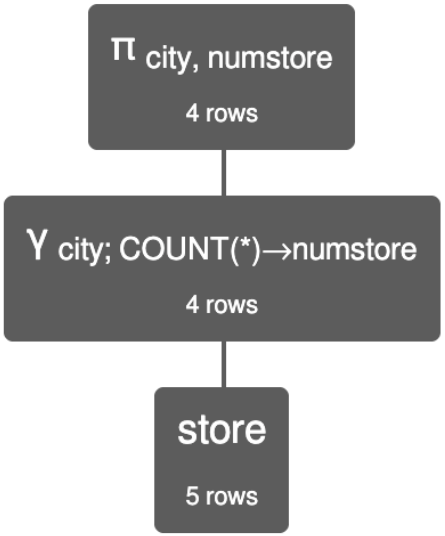




https://lbusuttil-um.github.io/database_analogies.html

https://lbusuttil-um.github.io/database_representation-based.html

```sql
select city, count(*) as numstore
from store
group by city
```



Strumenti?

store

Tempo di esecuzione: 0 ms

store.sID	store.sName	store.street	store.city
0	'Coop'	'Kalverstraat'	'Amsterdam'
1	'Lidl'	'Hoogstraat'	'Utrecht'
2	'Lidl'	'Molenstraat'	'Eindhoven'
3	'Hoogvliet'	'Rozemarijnstraat'	'Breda'
4	'Sligro'	'Stationsplein'	'Breda'

γ city; COUNT(*)→numstore store

Tempo di esecuzione: 2 ms

store.city	numstore
'Amsterdam'	1
'Utrecht'	1
'Eindhoven'	1
'Breda'	2

π city, numstore γ city; COUNT(*)→numstore store

Tempo di esecuzione: 2 ms

store.city	numstore
'Amsterdam'	1
'Utrecht'	1
'Eindhoven'	1
'Breda'	2

https://dbis-uibk.github.io/relax

Un possibile strumento: relaX

- Carica il dataset
  - Vai nel pannello Editor di gruppo
  - Copia e incolla il contenuto il tuo schema: **dbITADINFO.txt**
  - Clicca Anteprima per vederlo
  - Clicca Usa il gruppo nell'editor
- A questo punto vai nel pannello SQL e scrivi le tue query
- Puoi vedere il risultato parziale (contenuto di ogni nodo) durante l'esecuzione del comando cliccando sul nodo
- Ci sono molti dataset pronti

https://dbis-uibk.github.io/relax

Esercizi per chiarire?

Un esempio

- Colora in colori diversi le tuple corrispondenti a diversi valori di city nel db di esempio

- Cosa restituiscono (sull'esempio, in generale)

```
SELECT COUNT(*)                    e    SELECT COUNT(DISTINCT city)
FROM store;                             FROM store;
```

- Cosa restituiscono (sull'esempio, in generale)

```
SELECT COUNT(*)                    e    SELECT COUNT(DISTINCT sID)
FROM store                              FROM store
GROUP BY city;                          GROUP BY city;
```

A gruppi

Scegliete un errore frequente (preferibilmente in base alla vostra esperienza) e

1. Collocatelo nella tassonomia (primo livello)
2. Individuate possibili cause di quell'errore (convinzioni errate dello studente, origini della confusione)
3. Proponete
    1. un'analogia (con concetti del mondo reale o concetti matematici) che potrebbe aiutare a chiarirlo
    2. una visualizzazione che potrebbe aiutare a chiarirlo
    3. un esercizio mirato per lavorare su quell'errore

Link utili

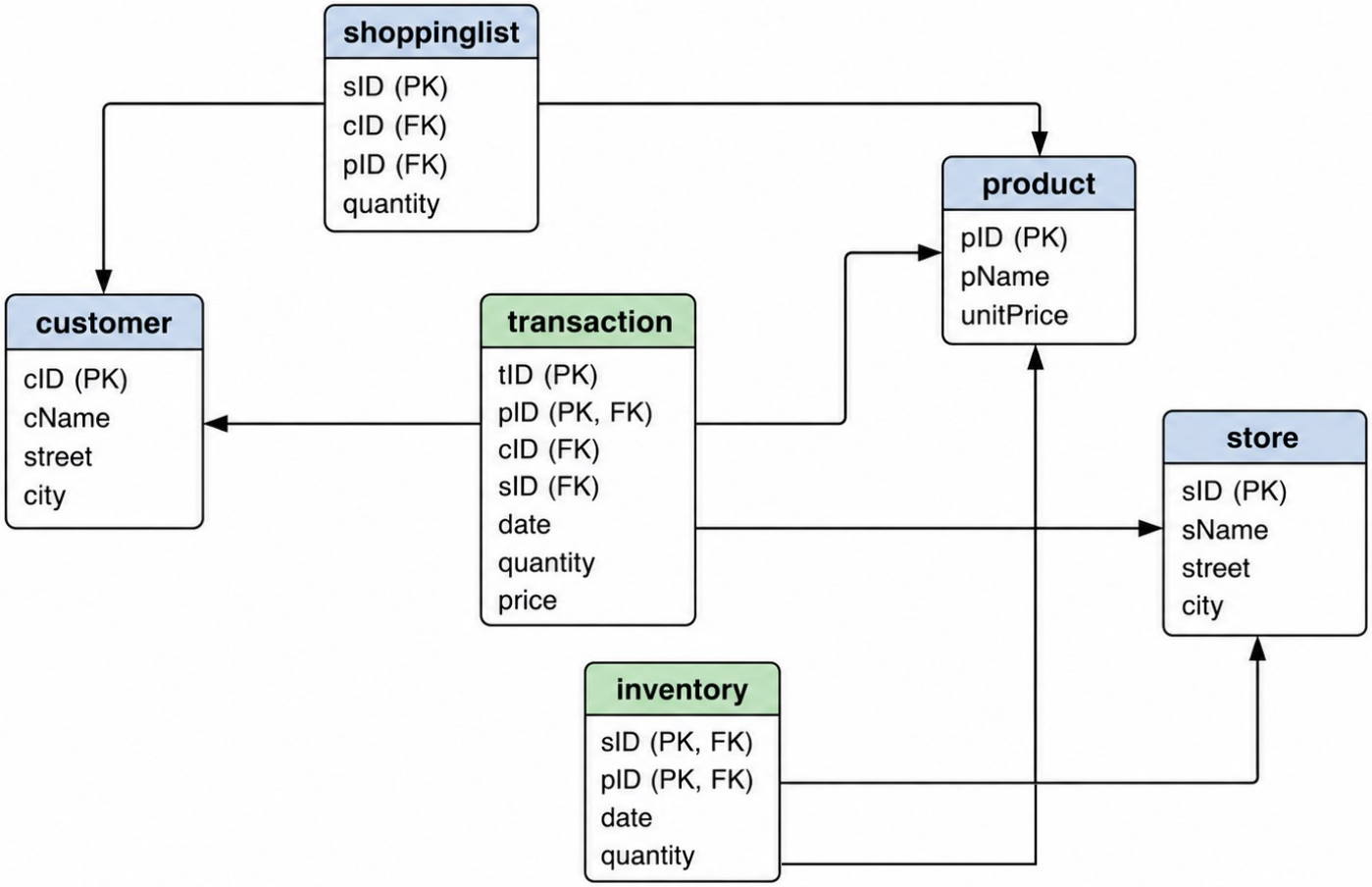
- Notional Machines for data systems
- https://daphnemiedema.nl/notionalmachines

- RelaX
- https://dbis-uibk.github.io/relax/landing

Lo schema è «neutro» rispetto agli errori nelle interrogazioni?

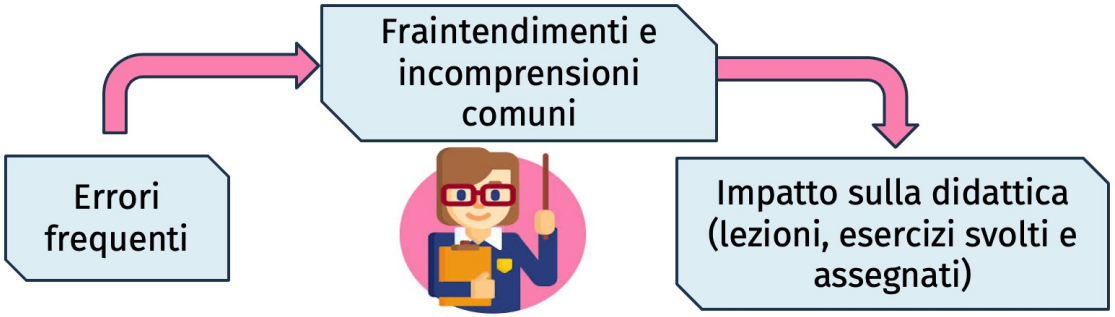
Cosa facilita?

Cosa può confondere?



e IA generativa

Motivazione didattica



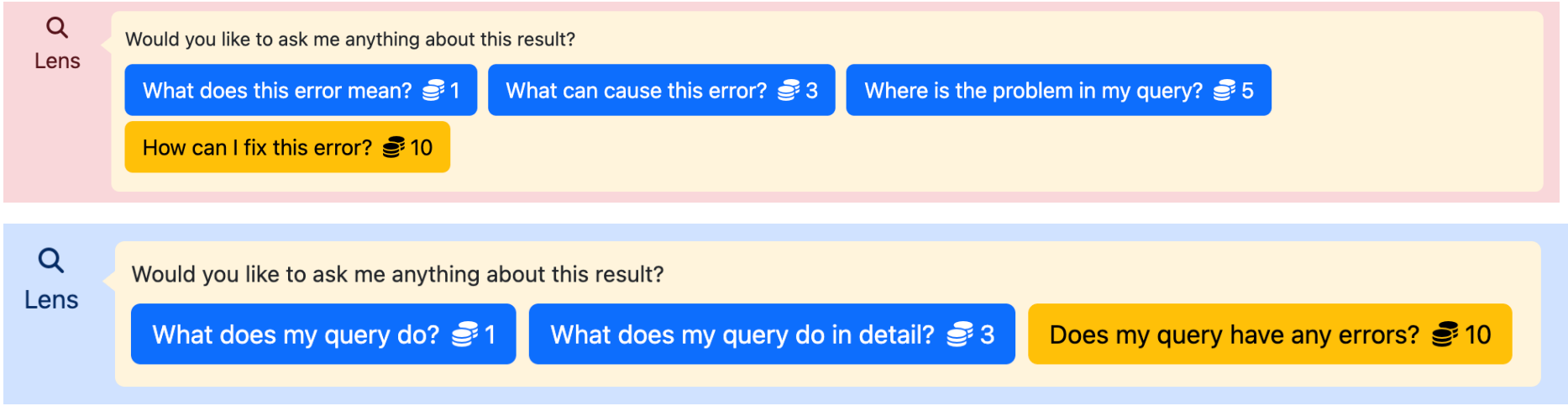
- Focus su errori frequenti e misconcezioni
- A granularità fine



- Meta-cognizione

LenSQL – genIA come tutor

LensQL



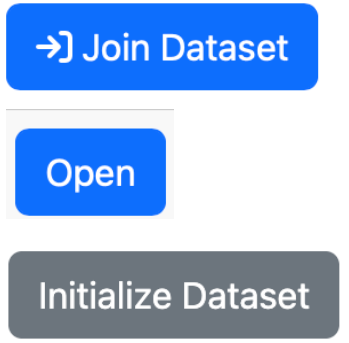
Ponzini, D., Catania, B., Guerrini, G. (2026). Enhancing SQL Learning Through Generative AI and Student Error Analysis. In: ADBIS 2025 https://doi.org/10.1007/978-3-032-05727-3_12

Proviamo

https://lensql.dibris.unige.it/

- Registratevi

- Unitevi al dataset con codice **EFA2GO4I**
- Aprite il dataset e un esercizio
- Inizializzate il dataset



- Provate a svolgere (alcune) interrogazioni sullo schema (minima variante di quello utilizzato prima) come se foste studenti alcune commettendo i loro errori tipici, alcune giuste
- Provate a utilizzare i vari bottoni

https://lensql.dibris.unige.it/ EFA2GO4I



Customer			
cID	cName	street	city
0	Noah	Koestraat	Utrecht
1	Sem	Rozemarijnstraat	Breda
2	Lucas	Oude Leliestraat	Amsterdam
3	Daan	Kalverstraat	Amsterdam

Shoppinglist			
cID	pID	quantity	date
1	2	1	2020-05-13
1	3	6	2020-05-13
3	1	2	2020-05-15

Store			
sID	sName	street	city
0	Coop	Kalverstraat	Amsterdam
1	Lidl	Hoogstraat	Utrecht
2	Lidl	Molenstraat	Eindhoven
3	Hoogvliet	Rozemarijnstraat	Breda
4	Sligro	Stationsplein	Breda

Inventory				
sID	pID	date	quantity	unit_price
0	1	2020-05-15	55	0.55
0	2	2020-05-15	32	2.3
1	4	2020-05-15	12	1.8
1	1	2020-05-15	46	0.6

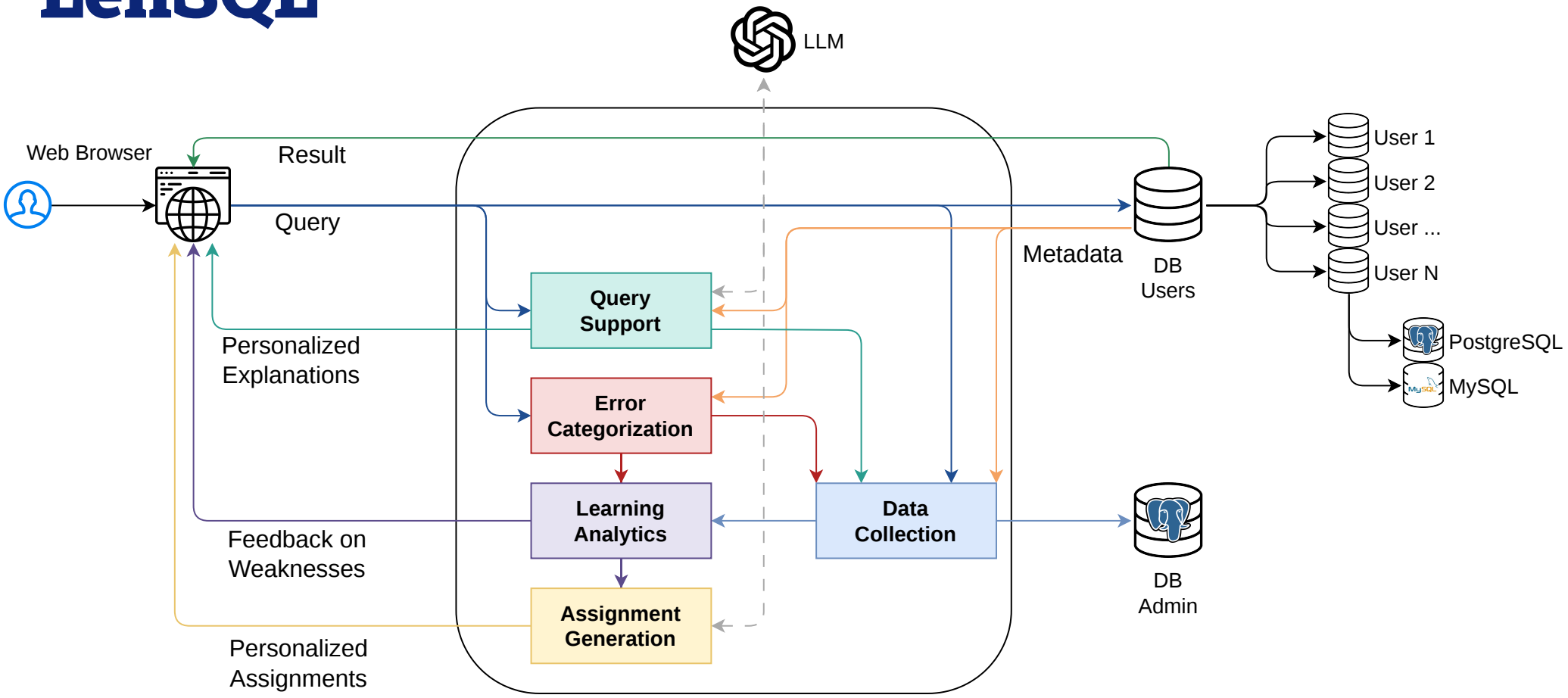
Transaction						
tID	cID	sID	pID	date	quantity	price
0	0	4	3	2020-05-12	5	0.4
1	0	4	1	2020-05-13	2	0.65
2	2	0	4	2020-05-13	2	1.3
3	3	0	1	2020-05-15	1	0.67

Product		
pID	pName	suffix
1	Milk	""
2	Mushrooms	""
3	Apples	""
4	Tea	""
5	Banana	""

Discussione

- Quali «bottoni» avete usato?
- La presenza di quel bottone è utile?
- Come ha funzionato?
  - Le risposte sono affidabili?
  - Le risposte sono utili?
- Ci sarebbero altri bottoni utili?
- Che prompt ci sono dietro?
- Guardiamo come è andata…

LenSQL


LLM
Web Browser
Result
Query
Personalized
Explanations
Feedback on
Weaknesses
Personalized
Assignments
Query
Support
Error
Categorization
Learning
Analytics
Assignment
Generation
Data
Collection
Metadata
DB
Users
DB
Admin
User 1
User 2
User ...
User N
PostgreSQL
MySQL

Supporto nell’esecuzione delle interrogazioni: feedback più comprensibile

LLM

Categorizzazione degli errori (adattamento di tassonomie in letteratura)

Attenzione studenti focalizzata sul processo di formulazione e sugli errori presenti nelle loro soluzioni

Affiancato ad attività di peer-review: verificare la presenza di errori tipici nelle soluzioni dei compagni (stessa tassonomia)

Generazione di esercizi personalizzati con un focus didattico sugli errori più frequenti

LLM

Usato nel nostro (e in altri) corsi di BD

Convergenza più rapida verso soluzione

reazione degli studenti positiva

- Grande apprezzamento del fatto che si parlasse apertamente di genAI e di come usarla

Ma …

- Un po’ di sfiducia rispetto a suggerimenti inutili
- Solo «inizio» della conversazione
- Condivisione dei prompt come esempi di possibili usi di genAI

Ringraziamenti & contatti

SQL e lenSQL

- Barbara Catania
- Hamid Livani, Davide Miggiano, Flavio Venturini, Elisa Calza, Angelo Guan, Musse Gebreezgabher, Shima Bagheeri
- Paola Podestà, Martin Goodfellow

giovanna.guerrini@unige.it

davide.ponzini@edu.unige.it

Se volete usare lenSQL con i vostri studenti contattateci!

Riferimenti

D. Miedema, E. Aivaloglou, G. Fletcher. Identifying SQL Misconceptions of Novices: Findings from a Think-Aloud Study. In Proc. ACM ICER 2021, pp. 355–367. https://doi.org/10.1145/3446871.3469759

D. E. Miedema. “On Learning SQL: Disentangling Concepts in Data Systems Education”. PhD thesis. Eindhoven University of Technology, 2024.

T. Taipalus, M. Siponen, and T. Vartiainen. “Errors and Complications in SQL Query Formulation”. In: ACM Transactions on Computing Education 18.3 (2018) . https://doi.org/10.1145/3231712

Davide Ponzini, Barbara Catania, Giovanna Guerrini. A Proposal for Revising SQL Error Taxonomies Based on Automated Detection. DataEd 2026

Daphne Miedema, Martin Goodfellow, Chandrika Satyavolu, Georgiana Haldeman, Leonard Busuttil, Laura Farinetti, Giovanna Guerrini, Yuhan Pan, Sujeeth Goud Ramagoni, Raja Sooriamurthi, and Xiaoying Tu. 2025. Explanations of Data Systems Concepts in CS Education: An Updated View on Notional Machines. In Proceedings of the 2025 Working Group Reports on Innovation and Technology in Computer Science Education. ACM, Nijmegen, Netherlands, 108–135.doi:10.1145/3760545.3783971

Ponzini, D., Catania, B., Guerrini, G. (2026). Enhancing SQL Learning Through Generative AI and Student Error Analysis. In: ADBIS 2025 https://doi.org/10.1007/978-3-032-05727-3_12

Barbara Catania, Giovanna Guerrini, Davide Ponzini, Elisa Calza (2026). Personalized SQL Exercise Generation with Large Language Models and Constraint-Based Validation. In: ADBIS 2026